Ed Yoon
Christina Baker
Mike Eisen
Pat Brown

Edward S. Yoon
Chief Technology Officer
Gordon and Betty Moore Foundation

February 25, 2002

Dear Mr. Yoon,

Thank you for considering our proposal. As you know, we represent the Public Library of Science, a non-profit organization of scientists whose goal is to make the world's scientific and medical literature a freely accessible resource for scientists, physicians and the public around the world.

We write to ask your support for a simple, specific plan that will believe will catalyze a revolutionary change in the distribution of scientific information and knowledge, and the ways it can be used.

We propose to establish an online scholarly publisher whose sole mission will be to make the original published reports of ideas, discoveries and research results in the life sciences and medicine (and eventually other fields) freely available online, without restrictions on use or further distribution, free from private or government control.

In parallel, we propose a plan to convert the most useful of the scientific and medical literature that has already been published into an openly accessible, public domain, digital resource.

If we succeed, a physician in a remote clinic in Ethiopia, a newly-diagnosed cancer patient in Salinas weighing her treatment options, or a budding scientist in a rural Mississippi high school, will be able to have, via the internet, the same access to the latest peer-reviewed scientific and medical research as a member of the faculty at Stanford or UCSF. Every scientist and physician will benefit from the ability to rapidly and comprehensively search and freely navigate the published record of scientific research and ideas, without artificial barriers or tolls.

Unrestricted access to published scientific information will also provide the essential foundation for development of diverse new ways to search, interlink and integrate the information in these published research reports – one of the most important frontiers in the life sciences. Scientists are eager to incorporate the information contained in research publications into their own databases (as they currently do with DNA sequences), to explore new ways to

integrate the contents of published works with information from disparate sources, to reorganize it, to annotate it, to map connections between pieces of information in disparate works published in different journals, and to transform it into something that goes far beyond an electronic version of journal volumes on a library shelf.

We believe that this project is directly responsive to the Moore Foundations goals of supporting university and graduate education in science and technology, and supporting scientific research. Moving the world's treasury of scientific and medical knowledge from books on the shelves of libraries in a few elite institutions to a freely accessible, searchable digital database would provide an enormous boost to research to all areas of science and medicine and increase opportunities for underrepresented minorities. Since this treasury of knowledge represents the product of hundreds of billions of dollars of public and private investment in research, the small additional investment required to make it an open digital resource would have extraordinary leverage. There are no good governmental sources of support for independent scientific publications.

There's a wonderful precedent that illustrates the "leverage" that you get just by making information openly accessible and useable by anyone - Genbank. Genbank was started about 20 years ago as a way to archive and distribute published DNA sequences - it was clear that they were best organized and analyzed using computers rather than ink on paper. The open access and unrestricted use of DNA sequence information, regardless of where it was published and with no obligation to the publisher, was absolutely essential in making possible the creative work of tens of thousands of scientists who have made that information immeasurably more useful than it would have been if had been treated in the way that most published scientific information is treated. The transformation of the life sciences by DNA sequences and genomics was absolutely dependent upon free and open access and unrestricted use of published DNA sequences - upon the ability to copy and use and transform and redistribute the information without any restrictions imposed by publishers. Imagine how much of the scientific progress of the past decade would have been sacrificed if the publishers had treated DNA sequences as they do all other published information. But DNA sequence information is only a small fraction of the published record of scientific research. The restrictions on access and use of the great majority of this information continue to stand in the way of an even greater creative transformation of science.

Why does this problem exist? Historically, scientists have relied on paper publication as the most efficient means for wide distribution and promotion of their work. When the information was encoded as ink on paper, and distributed using trucks and boats, a large fraction of the costs were in the printing and distribution, and each copy produced and distributed involved an expense for the publisher. The standard business model for scientific research publication, which organized works by scientific field into periodical journals sold by subscription, was sensible and efficient and served science and society well. The distribution of an author's work was limited only by the cost of printing and distributing copies. And the ability to find information in the huge body of published scientific work, or to map and record connections between bits of information published in separate works, in separate journals, was not limited by the business model, it was inherently limited by the physical nature of the paper literature - serial publications in physically dispersed volumes.

None of the essential premises of that business model remain valid today, as digital documents and electronic distribution have completely transformed the economics of distribution of information in science and medicine as in other fields. The costs involved in scientific publishing are almost entirely in steps leading to the original edited electronic document - the original is as expensive to produce as ever, but the costs to produce and distribute each additional copy electronically are now infinitesimal. Moreover, the profound limits that the paper format imposed on the ways that the information could be organized and accessed are now gone. The example of Genbank has demonstrated the extraordinary emergent value of freeing scientific information from that one-dimensional, one-paper-at-a-time organization.

We propose a simple alternative business model for scientific publication, designed to take full advantage of the economics and opportunities of electronic publication. In this model, the institutions that sponsor the research also pay the costs of publishing it, as an integral part of their mission to promote the discovery and dissemination of new knowledge. This cost, on the order of $1000 per published article, represents about 1% of the investment in the research itself. When an article is published, the authors grant to the public domain an irrevocable license to copy, hold, distribute, transform or otherwise use the

work. A faithful digital copy is provided to the National Library of Medicine for archiving, and digital copies are provided to multiple online distributors (such as PubMed Central), from which they can readily be accessed by scientists and the public.

The first immediate goal of our proposal is to launch a non-profit scientific publisher based on this business model (See Appendix A). Thanks to the support the Public Library of Science initiative (see: www.publiclibraryofscience.org) has already received from hundreds of the leading scientists in every field of biology and medicine, we are confident that we can assemble a group of editors and reviewers of unparalleled quality, to assure rigorous peer-review and rapid editorial processing of submitted works. Indeed, the scientists and physicians who are devoting their time and energy to this initiative include many past and present editors of prestigious scientific journals, leaders of scientific societies and major research institutions, and intellectual leaders in diverse disciplines.

We believe that we can successfully launch the proposed Public Library of Science publication system with an initial grant of $2 million, and continuing support at roughly $2 million annually for the first 5 years of operation. Eventually, we expect to recover all of our steady-state operating costs through modest charges to authors who can afford them (already a widespread practice for scientific journals).

The establishment of an open access digital archive of the scientific and medical research works already published will require a substantially larger investment (see appendix B), but even incremental investments will have enormous leverage.

Thank you for taking the time to consider this proposal.

Sincerely,

Patrick Brown, MD, PhD
Department of Biochemistry and
Howard Hughes Medical Institute
Stanford University School of Medicine

Michael Eisen, PhD
Lawrence Livermore Laboratories and
University of California, Berkeley

Harold Varmus, MD
President
Memorial Sloan-Kettering Cancer Center

on behalf of the Public Library of Science

Appendix A:

Public Library of Science Publications

January 31, 2002

Public Library of Science, a non-profit corporation, will publish scientific and medical research using a new business model that is intended to be financially sustainable and that results in the published work being freely available for use and distribution in the public domain. We aim not only to provide a high-quality vehicle for publication of scientific work, but also to establish the financial feasibility of this business model, so that it can be emulated by scientific societies that publish their own journals.

Our immediate goal is to launch a group of new journals that will:

1. Publish scientific and scholarly papers online.
2. Maintain high quality standards in editing and production.
3. Provide rigorous peer-review.
4. Archive the published papers as XML documents that conform to the PubMed Central DTD, so that conversion to new digital formats can be carried out by a single conversion script. (PMC has devoted a lot of effort to this archiving issue).
5. Distribute the published work to PubMed Central and other public distributors and repositories, including university libraries, etc.
6. Fund its operations by means of a combination of charges to authors, grant support from charitable organizations, Universities, and perhaps corporations who might benefit from freer access to the scientific literature.
7. Grant to the public domain an irrevocable license for unrestricted public distribution and use of the published work, requiring only proper citation of the original work.
8. Post an open record of all business operations, so that our efforts can serve as an experiment in testing a new business model for scientific publication.

The details of the organization and business plan will evolve over the next few months and further evolve as we gain experience. Although we will require "venture" funding for the start-up phase, we intend for this publishing operation to become financially self-sufficient.

Journal organization:

We propose initially to launch two online journals (tentatively: "Natural Sciences" and "Public Library of Medicine"), which will publish peer-reviewed reports that are deemed worthy of publication, but less "newsworthy" than the works that will be published in the top-tier journal. To make these journals as attractive and useful as possible to readers and potential authors, a substantial investment will be made to provide commentary and perspective pieces to accompany the original research articles, as well as short timely reviews, essays, meeting reports and news. This will require a significant investment above the bare-bones cost of electronic publication, but we believe that such an investment is critical to attracting authors and readers to a new publication of this kind.

Review and publication mechanism: Submissions will be electronic, and authors will be asked to prepare their manuscripts using a standard but flexible template (which we will provide at the PLoS website), to facilitate conversion of the submitted work to an XML document.

The peer review process will aim to: 1. decide whether the work is scientifically rigorous, intellectually honest, and presented and written clearly enough to be useful to its intended audience. (i.e. decide whether this article deserves to be published anywhere at all). 2. Decide what audience would benefit from reading the article, a judgment that currently would amount to deciding which journal it belongs in. In our model, it will be used to mark up the articles to facilitate creation of customized virtual journals for individual readers, based on their own interests and preferences.

We have already begun to organize a top-notch editorial group and a large group of committed reviewers, starting with the outstanding scientists who have signed the PLOS open letter, but also relying on their leadership to reach out to a much larger group of colleagues to recruit them as editors and reviewers. Our ability to draw on an outstanding group of highly respected scientists in diverse fields who share the goals of PloS will be a tremendous asset in establishing a prestigious brand identity and high visibility to a new journal.

A significant novel feature of these journals will increase their accessibility to a general audience – both the interested public reader, and scientific readers from outside of the scientific specialty directly addressed by a particular work. As an integral feature of the published work, the authors will be asked to submit with their manuscript a ca. 500 word (+ 1 or 2 simple illustration) piece that summarizes the work and explains its significance (from the authors' perspective) to the non-specialist - in language that an interested high-school or college student could understand (ie., roughly at the level of the NY Times' "science times"). This educational piece would be reviewed and edited with the same care as the primary article, with a view to ensuring that it is useful not only to make the scientific literature more accessible to the non-professional scientist, but also to help scientists browsing or searching for information.

The archival version of the published work, and the version of record - to which citations will point, will be the XML document residing at the National Library of Medicine, available from the PubMed Central sever, and numerous mirror sites and secondary distributors, as well as from the PLoS website itself. Each report will be rendered as an HTML document for direct viewing on the Web, and as a PDF for viewers who want to produce a paper copy. The Citation format will be, eg.: Public Library of Medicine 2(1): 1-10.

Despite the availability of free and unrestricted online access, experience and extensive discussion with our scientific colleagues shows that there is still a demand for printed versions of journals that publish a high density of newsworthy articles. We will therefore arrange for regular print publication of the contents of the PLOS journals, which we will provide to libraries and other institutions (and individuals) at a prices that cover the marginal cost of printing and distribution. Because the charges for the printed journals will not need to cover the cost of preparation of the original digital document (i.e. all the costs of the peer review, editing, and commissioned articles), the price of the printed journal should be very low, and attractive to many subscribers. This, in turn, will enhance the attractiveness of the PloS journals to prospective authors.

Draft Budget

Budget Category	*number*	*unit cost*	***total cost***
Expenses			
1. Personnel			
Managing Editor	1	$200,000	200000
Associate Editor	4	$90,000	360000
Business Manager	1	$90,000	90000
Programmer and IT person	1	$100,000	100000
Administrative Assistants	2	$50,000	100000
Copy editors	5	$60,000	300000
2. Consultants:			
Legal assistance		$20,000	20000
Business consultants		$50,000	50000
Graphic Designers		$10,000	10000
Accountant		$10,000	10000
3. Miscellaneous Fees			
Licenses		$2,000	2000
Telephone bills		$2,000	2000
Insurance		$20,000	20000
Utilities		$10,000	10000
4. Publishing services contract			
One-time start up cost		$100,000	100000
Annual fee		$100,000	100000
5. Equipment			
Computers/internet server/software		$40,000	40000
6. Office expenses			
Office space rental		$45,000	45000
Office furnishings/equipment		$40,000	40000
Office supplies and mail		$20,000	20000
7. Promotional costs		$10,000	10000
8. Travel		$20,000	20000
Subtotal fixed expenses			1649000
9. Publishing services (preparation of final edited original , maintenance of internet server)	4000	$300	1200000
Total expenses			-2849000
Income			
Grant support requested			2000000
Publication charges (assume 25% non-payment)	3000	$500	1500000
Balance			651000

Note on income and expenses that will scale with the success of the Public Library of Science project:

It is difficult to estimate how many reports we will publish in the first year of operation. Last year, more than 400,000 articles were published in scientific journals with a major focus on biology and medicine. If 1% of these articles were to be published by the Public Library of Science in a year of operation, then these fees would total $1,200,000.

We are presently planning to charge $500 per published article, and to waive this charge for authors who cannot afford to pay. If we assume that 20% of authors will ask for a waiver, then our total income from author fees will be $1,500,000, leaving a surplus of $1,200,000 to help defray fixed operating expenses. Because there will be an unpredictable but inevitable ramp-up phase, as well as unavoidable uncertainties in these financial projections and expected numbers of submissions, our budget builds in a surplus in the first year of $651,000. If indeed we have a surplus, we will use it to cover expansion costs and new projects and to buffer the budget in future years. If necessary, we will consider increasing the publication charges to $1,000 per article, to balance the budget.

Space: If we figure on needing around 2000 square feet of office space (reasonable?), near either Berkeley or Stanford, our rent is likely to be on the order of $2500-$3500/month.

Financial management: We propose to keep our financial records open, and post them online so that they can serve as a resource for other publishers who might wish to consider this business model. We will hire an independent accountant to audit our financial records on an annual basis and post the report online.

Although we believe that this enterprise will be extremely efficient and cost-effective, we are asking for a "venture funding" commitment of $10,000,000 over 5 years to insure a safe and successful launch, and to provide the assurance to the professional staff we will hire, and the authors and editors we recruit that the journal will have the resources to devote 5 years to establishing itself.

Appendix B: Transforming the archives of scientific and medical research into a public library of science

The goal:

We propose to bring together the world's published record of research discoveries, ideas, and knowledge in science and medicine, as recorded in scholarly journals and monographs, into a public-domain electronic public library, freely available online, at no charge and without restrictions of any kind, anywhere in the world. We will focus our initial efforts on the periodical literature in life sciences and medicine.

The published record of scientific and medical research represents the principal product of the work of hundreds of thousands of scientists and physicians. Much of the published record reports observations, ideas, and experimental results from work that was supported by public and charitable funds. The public investment in this effort over the past decade alone has exceeded 250 billion dollars. NIH alone will devote $20 Billion this year toward the advancement of medical knowledge.

A one-time investment to convert this treasury of information into an international public library will in effect provide a world-class scientific and medical research library to every university, college, high school, research institute, hospital, medical school, clinic, or public library that has an internet connection. Currently, only a minuscule fraction of the knowledge that has come from international scientific and medical research is readily available to the public or even to the scientific community.

Consolidating the published record into a single, unpartitioned, internationally-shared body of knowledge and ideas will also catalyze discovery and innovation, by making it possible to retrieve information much more efficiently, and to integrate the information into more rational structures than the current article-by-article, journal-by-journal format that is a vestige of the traditional library. Moreover, it is an essential step toward one of the next great challenges in biomedical research - the linking together and integration of the disparate fragments and islands of knowledge that have been separately gathered and published.

The cost of such a project is remarkably modest, when compared with cost of the research that produces the information contained in such a resource, or even the costs

of building a research library building or a research laboratory. A single visionary investment, on the order of $100,000,000, to build this new kind of library, will have more leverage, and more world-changing impact, than virtually any other investment in research or education, and its value will be realized in every corner of the world.

The Plan:

We propose to carry out this project in a staged, prioritized, systematic manner, focusing initially on the periodical literature in life sciences and medicine, and starting with the most important, recent and heavily used publications, and proceeding to the less important, older, and less heavily used. It's important to note that the 300 most heavily used journals account for less than 50% of all use of the published record of life sciences and medicine (as assessed by monitoring downloads of abstracts of articles through PubMed). Therefore, although we can set sensible priorities in this project, we should aim to be as comprehensive as possible in gathering the literature into this public resource.

There are three components to the plan:

1. Fund Raising:

We expect that the great potential benefits to biomedical science and public health will motivate public-spirited individuals, charitable organizations - (particularly those with an interest in research, public health, education and libraries), universities, hospitals and health-care providers, pharmaceutical and biotechnology companies, and computer, internet, and information technology companies to join forces to finance this initiative. A founding investment by a leading philanthropist is critical to bringing together this larger community of interests.

2. Mechanics and timeline of digitization, archiving and distribution.

We will proceed in a deliberate, prioritized manner, beginning with the journals and volumes that are most widely and heavily used. It is important to note, however, that although a relatively small number of journals are disproportionately read and cited, these high-impact journals still account for only a small fraction of the published record that scientists and physicians rely on. For example, the 300 journals whose article abstracts were

most frequently requested in PubMed searches accounted for fewer than half of all such requests. Similarly, although recent journals were disproportionately used, more than half of all abstract hits were for articles more than 2 years old, and a quarter were for articles older than 10 years old. Therefore we propose to identify, based on citation frequency, the frequency of abstract requests in PubMed searches, and information from Medical School librarians on library usage patterns, a prioritized list of journals, by year, for digitization and incorporation into the Public library.

The second important element of the plan is the format and quality of the digitized material. For many journals, digital-format versions of articles published in the past few years are already available. For the remainder, we propose to begin by converting journal pages to 600 dpi b&w TIFF images, and Figures to 200 dpi color or gray-scale JPEG images. These high-resolution images of the original journal pages will, at least initially, be the primary format for providing the material electronically to readers. OCR'd ASCII files of the text will be used to provide full text searching and indexing of the contents, and the frontmatter and backmatter and citations (that is, the title, authors, affiliations, abstract, received date, publication date, journal, volume, issue, page, and all citations) will be converted to fully-tagged XML files with very high accuracy. Since this level of digital conversion has been used previously by others (eg. the Americal Physical Society), and we have a tentative price quotation from the company that has handled this process for the American Physical Society, we can make a reasonable estimate of the cost, throughput and quality for the digitization process.

Ideally, the digitization process will proceed in a staged fashion as follows. (In practice, it will have to be modified based on the responses of the copyright owners).

Phase I (2 years?): Approximately 3,000 journal-years; the most recent 10 years of 300 most widely used journals converted to digital format as above. Preliminary studies indicate that this set will account for approximately 40% of the total use of the archival literature. (Certainly it will provide a resource considerably more complete than almost any departmental library). This will amount to about 1.5 million articles, and about 9 million pages, for an estimated cost of $11-20 million for the digitization

process itself. (Scanning and OCR - between 0.30 and 1.00/pg, depending on level of editing and QC of the OCR= $2.7-9M; images of figures, 0.60/figure x estimated 5 figures/article = $4.5M; XML tagging of front and backmatter, at 0.90/kilobyte x estimated 3-4 kb/article=$4-5M. The XML costs are likely to be lower because much of the frontmatter has already been tagged by NLM for Medline). The additional cost of obtaining the rights (see below) and the physical copies of the journals to be digitized, and the cost of packing and shipping to the contractors (in India) are likely to be at least several million dollars more.

In addition, we will need at the earliest stage to provide servers for storage and distribution, a staff to manage the ongoing program and to maintain the servers and database, and an endowment for preserving archival copies of the digitized material and ensuring their preservation. We estimate that at the end of Phase I, the archive will require about 3TB of disk storage (about 300 kB per page) which would cost about $300,000, and at the end of Phase II (see below), it would require about 10 TB of disk storage, costing around $500K to $1M. We intend to distribute the entire archive on an ongoing basis to mirror sites and to individual institutions throughout the world, so there would be a similar cost to those institutions for initial purchase of a server with the needed storage capacity.

Phase II (2 years?): Digitize the previous 10 years of top 300 journals, plus most recent 10 years of journals ranked between 301-2000. Preliminary studies indicate that the material digitized in Phases I and II will account for about 80% of the total use of the archival literature. Because the less-highly used journals, and older journals, tend to have fewer pages per year, this phase will amount to about 5 million articles and 30 million pages. The estimated cost will therefore be about $38-63 million for the digitization process itself.

Phase III (continuing): Continue the process through the less heavily-used literature. This is an open-ended process, and the plan can evolve as we learn from the results of Phase I and Phase II, and from user feedback, what features and what materials are of highest priority, and as the available budget for ongoing activities becomes more clear.

3. Obtaining the rights for unrestricted public distribution of previously-published works. The strategy has two stages: First, ask publishers to donate copyrights, or a license for unlimited distribution of the published material whose copyrights they currently own. Preliminary discussions with many scientific society and university publishers have shown that most scientific societies are likely to support this endeavor. In addition, preliminary discussions with some for-profit publishers (eg., Cold Spring Harbor Press) suggests that even many commercial publishers are likely to be willing to donate, or sell at a modest cost, rights for archival material that is producing little or no current revenue. Second, in some cases, publishers are likely to ask for some monetary compensation for the material they provide. We should offer to pay fair compensation as a single one-time purchase price for an unrestricted permanent license (eg. Estimate the revenue they currently receive from sales of archival issues/reprints, as a function of age of the material, and integrate to infinity). We should emphasize that as this electronic public library grows, it is likely to become a highly-used, highly visible resource, and journals whose archival material is withheld will lose some of their visibility and presence. More importantly, it will diminish the attractiveness of the non-participating journals to the authors of new works.

Clearly, the cooperation of the publishers who own the copyrights in the archival works will have a major role in determining the course of this project.